

Pengelompokan Data Kajicuaca Menggunakan Teknik Peraturan Kesatuan

Rozilawati Dollah @ Md. Zain¹, Mohd Noor Md Sap²,
Mahadi Bahari³, Aryati Bakri⁴
Jabatan Sistem Maklumat
Fakulti Sains Komputer & Sistem Maklumat
Universiti Teknologi Malaysia
81310 Skudai, Johor

{zeela¹,mohdnoor²,mahadi³,aryati⁴@fsksm.utm.my}

Abstrak

Penganalisan data merupakan satu tugas yang penting bagi pihak Jabatan Perkhidmatan Kajicuaca Malaysia (JPKM), di mana ianya digunakan untuk membuat peramalan taburan hujan pada masa akan datang. Pengekstrakan maklumat yang berguna daripada pangkalan data di JPKM bagi tujuan penganalisan ini menjadi sukar ekoran pertambahan jumlah data kajicuaca yang disimpan. Peramalan taburan hujan boleh diaplikasikan melalui operasi perlombongan data iaitu pengelompokan data. Kajian ini memberikan penumpuan kepada operasi pengelompokan data kajicuaca di mana hasil dari pengelompokan ini akan digunakan untuk melakukan peramalan taburan hujan. Hasil kajian ini menunjukkan pengelompokan data kajicuaca dengan menggunakan teknik peraturan kesatuan dapat membantu pihak peramal di JPKM untuk melakukan peramalan taburan hujan. Selain daripada itu, hasil kajian juga mendapati bahawa penggunaan kelompok data kajicuaca yang berlainan sebagai data input kepada proses peramalan taburan hujan telah memberikan prestasi peramalan yang lebih baik berbanding dengan penggunaan kelompok data kajicuaca yang sama.

Katakunci: perlombongan data, pengelompokan, penganalisan data, peraturan kesatuan dan peramalan taburan hujan

1.0 Pengenalan

Jabatan Perkhidmatan Kajicuaca Malaysia (JPKM) memainkan peranan yang penting di dalam memantau situasi perubahan cuaca dan mengeluarkan kenyataan, nasihat dan amaran cuaca bila keadaan memerlukan (JPKM, 2000). Ini bermakna, JPKM memerlukan maklumat data kajicuaca yang tepat dari setiap stesenya untuk membuat jangkaan atau peramalan yang terperinci tentang perubahan cuaca sebelum ianya dihebahkan. Antara parameter-parameter data kajicuaca ialah data taburan hujan, suhu, angin, kelembapan, sejatan dan sinaran suria (Sani, 1984). Peningkatan jumlah data ini telah menimbulkan kesukaran kepada pihak JPKM untuk melakukan proses penganalisan data kajicuaca termasuklah bagi tujuan melakukan peramalan cuaca.

Di antara parameter-parameter kajicuaca, taburan hujan merupakan parameter yang paling sukar diramal (Liu dan Lee, 1999) memandangkan peramalan hujan merupakan satu masalah yang kompleks dan sukar kerana ia melibatkan pelbagai pembolehubah di mana ianya saling berhubungkait dengan cara yang rumit. Kebanyakan perhubungannya adalah menggambarkan hubungan ruang dan dinamik yang tidak linear (Chen dan Takagi, 1993). Sehubungan dengan itu, data taburan hujan telah dipilih untuk diramal di dalam kajian ini.

Berdasarkan kepada kajian tentang peramalan atau penganggaran taburan hujan yang telah dijalankan oleh beberapa penyelidik sebelum ini, didapati kebanyakan kajian, contohnya

kajian yang dijalankan oleh Chen dan Takagi (1993), McCullagh et. al (1999) serta Liu dan Lee (1999) telah menggunakan teknik perlombongan data rangkaian neural pintar (*Artificial Neural Network*) untuk melakukan operasi pengkelasan terhadap data taburan hujan tersebut bagi tujuan peramalan atau penganggaran taburan hujan. Walaubagaimanapun, masih terdapat kekurangan di dalam kajian mereka yang antaranya ialah berlakunya *overlappes* di dalam set data semasa proses latihan. Sehubungan dengan itu, kajian yang mereka jalankan ini adalah tidak sesuai untuk digunakan bagi data input yang terlalu banyak.

Oleh yang demikian, kajian ini dilaksanakan untuk membantu pihak JPKM melakukan penganalisaan data kajicuaca bagi tujuan meramal taburan hujan dengan penumpuan kepada operasi pengelompokan data kajicuaca di mana hasil dari pengelompokan ini akan digunakan untuk melakukan peramalan taburan hujan. Operasi pengelompokan data kajicuaca bagi tujuan peramalan taburan hujan telah dipilih di dalam kajian ini kerana terdapat banyak kajian yang melibatkan pengelompokan telah berjaya meningkatkan hasil peramalan yang dilakukan. Contohnya, kajian yang dijalankan oleh Sarjon dan Mohd Noor (2000), mendapati pengelompokan dapat meningkatkan peramalan yang dilakukan. Operasi pengelompokan ke atas data kajicuaca ini akan diperjelaskan dengan terperinci di dalam seksyen dua nanti. Bagi melaksanakan kajian ini, teknik pengelompokan peraturan kesatuan telah dipilih untuk melakukan pengelompokan data kajicuaca. Oleh itu, penulis akan membincangkan tentang konsep teknik ini di dalam seksyen tiga nanti. Seksyen empat akan meliputi metodologi yang digunakan di dalam kajian ini. Seksyen lima pula akan mengupas mengenai beberapa siri eksperimen dan analisis keputusan yang dijalankan manakala kajian lanjutan yang akan datang diperbincangkan dalam seksyen enam. Kertas kerja ini akan diakhiri dengan kesimpulan bagi tajuk yang telah dibincangkan.

2.0 Pengelompokan

Pengelompokan juga dikenali sebagai pengsegmenan di mana ia tidak menghasilkan medan yang spesifik untuk diramal tetapi ia menentukan sasaran data kepada beberapa subset yang sama dengan medan lain. Menurut Fayyad (1997), perangkaian ialah proses mengumpulkan objek fizikal atau abstrak yang sama ke dalam kelas-kelas untuk mengetahui corak agihan keseluruhan set-set data.

Perangkaian juga boleh ditakrifkan sebagai sekumpulan objek atau data yang dirangkaikan mengikut kriteria sama. Ia selalunya dicapai dengan mencari sekumpulan perkara data yang hampir antara satu sama lain berdasarkan kepada beberapa kriteria. Menurut TwoCrows (1999) pula, pengelompokan ialah membahagikan pangkalan data kepada beberapa kumpulan. Matlamat bagi pengelompokan ialah untuk memperolehi kumpulan yang berbeza di antara satu sama lain tetapi ahli bagi setiap kumpulan tersebut adalah hampir sama. Manakala Kaski (1997) pula mentakrifkan pengelompokan merupakan kaedah untuk menghimpunkan vektor-vektor yang hampir sama atau yang mempunyai persamaan antara satu sama lain berdasarkan ciri-ciri datanya. Ia seperti merangkaikan atau menghimpunkan data-data berdasarkan persamaan ciri data tersebut. Daripada beberapa definisi yang dinyatakan di atas, maka secara keseluruhannya dapatlah dikatakan bahawa pengelompokan merupakan proses carian terhadap corak dalam pangkalan data bersejarah untuk mengelompokkan data-data yang mempunyai persamaan ke dalam satu kelompok atau kumpulan.

Pengelompokan merupakan salah satu daripada masalah utama di dalam penjelajahan analisis data. Masalah pengelompokan data berlaku di dalam pengecaman corak, statistik, pembelajaran tanpa seliaan, rangkaian neural, perlombongan data, pembelajaran mesin dan pelbagai bidang saintifik yang lain (Hofmann dan Buhmann, 1997). Operasi pengelompokan juga agak popular di kalangan penyelidik dewasa ini. Banyak kajian yang telah melibatkan pengelompokan dijalankan. Contohnya kajian yang dijalankan oleh Sarjon dan Mohd Noor

(2000a) mendapati, penggunaan pengelompokan dapat meningkatkan prestasi peramalan yang dilakukan. Selain daripada itu, kajian oleh Yair, et. al (1999) juga menunjukkan bahawa penggunaan pengelompokan semantik berdasarkan metrik persamaan semantik dapat meningkatkan hasil peramalan yang dilakukan. Kajian pengelompokan juga telah dijalankan oleh Sakira (1998); Azah (1999) dan Shahliza (1999).

3.0 Peraturan kesatuan

Peraturan kesatuan merupakan salah satu daripada teknik-teknik perlombongan data di mana ia digunakan untuk menemukan kesatuan yang kuat atau kukuh dan juga mencari perhubungan di antara data (Chen, 1999). Peraturan yang ditemui akan membantu di dalam pemasaran atau analisis jualan, membuat keputusan dan pengurusan perniagaan. Ordenez telah menggunakan peraturan kesatuan untuk meramalkan penyakit jantung dengan menggunakan data perubatan (Ordenez, etc, 2001). Peraturan kesatuan digunakan untuk tujuan pergabungan semulajadi di atas data-data perubatan. Peraturan kesatuan digunakan untuk mencapai matlamat perlombongan data termasuklah permodelan data, peramalan pendapatan masa hadapan dan juga untuk menyokong dalam membuat keputusan (Choenni dan Siebes, 1996).

Teknik ini menumpukan kepada pangkalan data bertransaksi atau berhubungan. Matlamat utama teknik peraturan kesatuan ini ialah untuk mencari perkaitan di antara koleksi item-item pada satu set transaksi. Perkaitan di antara item-item inilah diwakili oleh peraturan kesatuan. Prosidur asas bagi perlombongan peraturan kesatuan mengandungi dua langkah:

- i) Mengasingkan set-set data yang besar daripada pangkalan data bertransaksi,
- ii) Hasilkan peraturan kesatuan daripada set-set yang besar itu.

Di dalam teknik ini, dua parameter yang digunakan untuk menentukan kesahihan dan keutamaan peraturan di antara item-item yang dikaitkan. Penentuan atau pengukuran peraturan ini dibahagikan kepada dua iaitu *support* (s) dan *confidence* (c).

Bentuk Peraturan : “Badan -> Kepala [*support*, *confidence*]”

Faktor *support* adalah parameter yang menunjukkan peraturan kesatuan yang telah dikesan dari keseluruhan set data di dalam satu transaksi. Manakala faktor *confidence* pula merupakan darjah yang mengatakan bahawa peraturan tersebut adalah betul bagi setiap rekod. Nilainya adalah dikira seperti berikut (Fu, 1996 ; Chen, 1999) :

$$Support (X \Rightarrow Y) = \text{Kebarangkalian } (X \cup Y) \quad \dots(a)$$

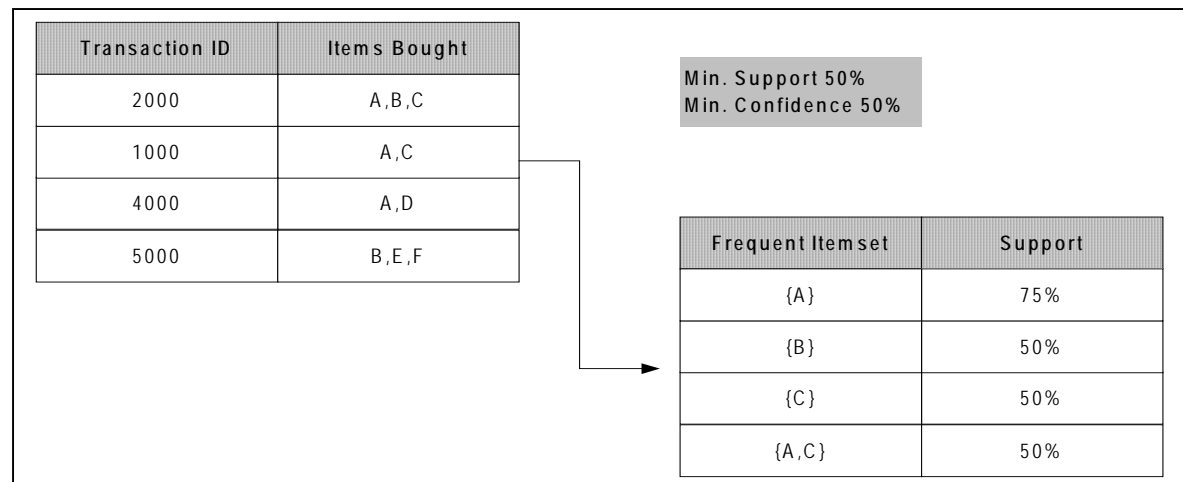
$$Confidence (X \Rightarrow Y) = \text{Kebarangkalian } (Y / X) \quad \dots(b)$$

Di mana, kebarangkalian $(X \cup Y)$ merupakan kebarangkalian kedua-dua X dan Y berlaku di dalam pangkalan data dan kebarangkalian (Y / X) pula merupakan kebarangkalian bersyarat bahawa Y berlaku bila X berlaku. Oleh yang demikian, *support* mengukur kekerapan bagi corak kejadian dan *confidence* pula mengukur kekuatan dakwaan tersebut.

Sebelum melombong perkaitan di antara item-item di dalam pangkalan data, ID bagi transaksi disusun secara berturutan di dalam sumber input. Setiap transaksi adalah senarai item-item. Teknik ini bermula dengan mengira kejadian setiap item yang ada di dalam set data bersama dengan transaksi dan membina satu dimensi lajur atau vektor di mana di dalam setiap sel

mengandung kiraan bagi setiap satu item. Jika kiraan di dalam sel adalah di bawah atau kurang dari aras *support*, maka sel tersebut akan diabaikan.

Seterusnya dua dimensi matriks dibentuk untuk menyimpan kiraan berlakunya setiap item dan setiap sel akan ditapis jika *support* kurang dari nilai *threshold*. *Threshold* adalah nilai minimum yang diberi untuk setiap parameter iaitu *support* dan *confidence*. Contoh bagi perlombongan peraturan kesatuan adalah seperti yang ditunjukkan oleh Rajah 1 di bawah;



Rajah 1 : Contoh perlombongan peraturan kesatuan

Bagi peraturan $A \rightarrow C$:

$$Support(A \rightarrow C) = \frac{Support(\{A,C\})}{Support(A)} = 50\%$$

$$Confidence(A \rightarrow C) = Support(\{A,C\}) / Support(\{A\}) = 66.6\%$$

Jika dilihat, nilai bagi *support* dan *confidence* melebihi nilai *threshold* yang diberi. Jika nilai *support* dan *confidence* bagi sesuatu peraturan adalah tinggi maka ia menunjukkan darjah kesahihan yang tinggi bagi peraturan yang diperolehi berbanding dengan peraturan yang mempunyai nilai *support* dan *confidence* yang rendah. Untuk menjanakan set item yang frequent ini, algoritma Apriori yang diperkenalkan oleh Agrawal dan Srikant merupakan algoritma yang popular dan banyak digunakan oleh penyelidik-penyelidik.

4.0 Metodologi

Perlaksanaan kajian ini adalah melibatkan beberapa aktiviti. Antara aktiviti-aktiviti yang terlibat di dalam perlaksanaan kajian ini adalah seperti berikut :-

- i) **Mengumpul dan menganalisa data kajicuaca** : data yang digunakan di dalam kajian ini adalah terdiri daripada data kajicuaca (setiap jam) dari bulan September 1993 hingga bulan Februari 2001, yang diperolehi daripada stesen penceraian Kluang (MPOB). Jadual 1 di bawah menunjukkan sampel data kajicuaca. Di antara atribut data kajicuaca yang digunakan dalam kajian ini ialah *windvane*, kelembapan, *energy*, suhu, tekanan, *radiation*, kelajuan angin dan taburan hujan.

Jadual 1 : Sampel data kajicuaca

ID	Tarikh	Masa	Windvane	Kelembapan	Taburan Hujan
1	8/1/00	0:00	197.0	0.0		0.0
2	8/1/00	1:00	201.2	0.0		0.0
3	8/1/00	2:00	206.5	0.0		0.0
...
...
250	8/11/00	9:00	222.5	0.0		0.0

- ii) **Pengkelasan data kajicuaca** : sebelum data kajicuaca dikelompokkan, ia perlu dikelaskan terlebih dahulu. Tujuan pengkelasan data kajicuaca ini dilakukan ialah untuk memudahkan pengiraan nilai *support* dan *confidence*. Bagi tujuan ini, kaedah kuantitatif yang diperkenalkan oleh Fact, (1987) telah digunakan untuk membina pengagihan kelas. Senarai kelas-kelas yang telah dibina boleh dirujuk pada Lampiran A. Jadual 2 di bawah menunjukkan data kajicuaca yang telah dikelaskan.

Jadual 2 : Sampel data kajicuaca yang dikelaskan

ID	Tarikh	Masa	Windvane	Kelembapan	Taburan Hujan
1	8/1/00	0:00	W1	H1		RF1
2	8/1/00	1:00	W2	H1		RF1
3	8/1/00	2:00	W2	H1		RF1
...
...
250	8/11/00	9:00	W2	H1		RF1

Di mana ; W = *Windvane*, H = *Kelembapan*, E = *Energy*
 T = *Suhu*, TE = *Tekanan*, R = *Radiation*,
 WS = *Kelajuan Angin*, dan RF = *Taburan Hujan*.

- iii) **Mengira nilai support dan confidence bagi data kajicuaca** : berdasarkan kepada sampel data kajicuaca dalam Jadual 2 di atas, pengiraan nilai *support* dan *confidence* dilakukan ke atas data tersebut. Pengiraan nilai *support* dan *confidence* bagi data-data kajicuaca ini dilakukan bertujuan untuk mencari set item yang *frequent*. Ia dilakukan dengan menggunakan teknik peraturan kesatuan. Teknik peraturan kesatuan ini digunakan untuk mencari item (atribut) yang dikatakan mempunyai perkaitan dengan item (atribut) yang lain dalam satu pangkalan data. Faktor *support* adalah parameter yang menunjukkan peraturan kesatuan yang telah dikesan dari keseluruhan set data di dalam satu transaksi.

Manakala faktor *confidence* pula merupakan darjah yang mengatakan bahawa peraturan tersebut adalah betul bagi setiap rekod. Nilainya adalah dikira seperti berikut :

$$\text{Support } (X \Rightarrow Y) = \text{Kebarangkalian } (X \cup Y)$$

$$\text{Confidence } (X \Rightarrow Y) = \text{Kebarangkalian } (Y / X)$$

Di mana;

Kebarangkalian $(X \cup Y)$ merupakan kebarangkalian kedua-dua X dan Y berlaku di dalam pangkalan data dan kebarangkalian (Y / X) pula merupakan kebarangkalian bersyarat bahawa Y berlaku bila X berlaku.

- iv) **Mendapatkan set item yang frequent** : nilai *support* dan *confidence* yang diperolehi akan digunakan untuk mengasingkan set item yang *frequent* dan set item yang tidak *frequent*. Ia ditentukan dengan membandingkan nilai *support* dan *confidence* yang diperolehi dengan nilai minimum *support* dan juga nilai minimum *confidence*. Nilai minimum *support* dan nilai minimum *confidence* yang digunakan di dalam kajian ini ialah 0.1. Oleh yang demikian, bagi data kajicuaca yang melebihi atau sama dengan nilai minimum *support* dan minimum *confidence*, maka ia akan menjadi atribut atau set item yang *frequent*. Selain daripada itu, salah satu cara yang boleh digunakan untuk menjanakan set item yang *frequent* ini ialah algoritma Apriori yang telah diperkenalkan oleh Agrawal dan Srikant (Sarjon dan Mohd Noor, 2000).
- v) **Mengelompokkan set item yang frequent** : selepas dibuat perbandingan di antara nilai *support* dan *confidence* dengan nilai minimum masing-masing, set item yang *frequent* akan diperolehi. Atribut atau set item yang *frequent* ini kemudiannya dikelompokkan ke dalam kelompok atau kumpulan yang sama manakala set item yang tidak *frequent* pula dikelompokkan ke dalam satu kelompok yang berlainan. Bagi tujuan pengujian teknik pengelompokan ini, hanya kelompok set item yang *frequent* sahaja yang akan digunakan sebagai data input kepada proses peramalan taburan hujan.
- vi) **Peramalan taburan hujan** : kelompok-kelompok set item yang *frequent* yang telah dihasilkan ini kemudiannya digunakan sebagai data input kepada Pakej NeuNetPro 2.3 bagi tujuan meramal taburan hujan. Sehubungan dengan itu, beberapa eksperimen yang melibatkan kelompok-kelompok data kajicuaca yang sama dan juga kelompok-kelompok data kajicuaca yang berlainan telah dilaksanakan. Ianya bertujuan untuk membuat perbandingan prestasi peramalan taburan hujan di antara kelompok data kajicuaca yang sama dan juga kelompok data kajicuaca yang berlainan.
- vii) **Penganalisaan dan perbandingan hasil peramalan** : keputusan peramalan taburan hujan telah digunakan untuk membuat penganalisaan dan perbandingan di antara kelompok data kajicuaca yang sama dengan kelompok data kajicuaca yang berlainan. Ianya bertujuan untuk melihat sejauhmanakah ketepatan dan keberkesanan di antara teknik peraturan kesatuan di dalam mengelompokkan data kajicuaca. Di samping itu, ianya juga bertujuan untuk melihat kesan penggunaan kelompok data kajicuaca manakah yang memberikan pengaruh kepada prestasi peramalan taburan hujan. Perbandingan ini dilakukan dengan mencari perbezaan di antara nilai taburan hujan yang sebenar dan nilai taburan hujan yang diramal. Selain daripada itu, nilai ralat min punca kuasa dua (RMS) dan pekali korelasi juga turut digunakan bagi tujuan perbandingan keberkesanan dan ketepatan peramalan.

5.0 Eksperimen

Di dalam kajian ini, pengelompokan data kajicuaca telah dilaksanakan dengan menggunakan salah satu teknik perlombongan data, iaitu teknik peraturan kesatuan. Teknik peraturan kesatuan telah dipilih kerana kebolehanannya di dalam menemukan kesatuan yang kukuh atau kuat dan juga mencari perhubungan di antara koleksi data-data. Selain daripada itu, teknik peraturan kesatuan juga merupakan salah satu kaedah penganalisan data, di mana ia berupaya membina model peramalan dari jumlah data yang besar.

Pengelompokan bermaksud membahagikan satu pangkalan data kepada beberapa kelompok yang mempunyai ciri atau sifat yang sama. Apabila atribut data yang mempunyai ciri yang sama telah dikumpulkan ke dalam satu kelompok, ini bermakna atribut data yang berada di dalam kelompok yang lain mempunyai ciri yang berbeza daripada mereka. Atribut-atribut di dalam satu kelompok mestilah mempunyai persamaan yang boleh diukur menggunakan kaedah tertentu, contohnya mengira jarak di antara atribut-atribut tersebut.

Terdapat pelbagai cara untuk mengelompokkan data kajicuaca di dalam satu pangkalan data. Di antaranya ialah dengan menggunakan kaedah statistik, teknik perlombongan data dan sebagainya. Sehubungan dengan itu, kajian ini telah dilaksanakan dengan menggunakan teknik peraturan kesatuan, iaitu salah satu teknik dalam perlombongan data. Jadual 3 di bawah menunjukkan sampel data kajicuaca yang telah dikelaskan di dalam kajian ini.

Jadual 3: Sampel data kajicuaca yang dikelaskan

ID	Tarikh	Masa	Windvane	Kelembapan	Taburan Hujan
1	8/1/00	0:00	W1	H1		RF1
2	8/1/00	1:00	W2	H1		RF1
3	8/1/00	2:00	W2	H1		RF1
...
...
250	8/11/00	9:00	W2	H1		RF1

Untuk melaksanakan pengelompokan data kajicuaca, teknik peraturan kesatuan akan mencari perkaitan di antara koleksi item-item atau atribut-atribut dalam pangkalan data kajicuaca. Perkaitan ini diperolehi dengan cara mencari set item yang *frequent*. Ia dilakukan dengan mengira nilai *support* dan juga nilai *confidence* bagi set item atau atribut yang terdapat di dalam pangkalan data kajicuaca. Faktor *support* dan *confidence* merupakan dua parameter yang digunakan untuk menentukan kesahihan dan keutamaan peraturan di antara atribut-atribut yang dikaitkan.

Sebelum data kajicuaca dikelompokkan, ia perlu kelaskan terlebih dahulu. Tujuan pengkelasan ini dilakukan ialah untuk memudahkan pengiraan nilai *support* dan *confidence* dilakukan. Berdasarkan kepada jadual sampel data kajicuaca di atas, pengiraan nilai *support* dan *confidence* dilakukan ke atas data tersebut. Hasil daripada pengiraan nilai *support* dan *confidence* yang diperolehi bagi setiap atribut data kajicuaca yang digunakan di dalam kajian ini ditunjukkan oleh Jadual 4;

Jadual 4 : Nilai *Support* dan *Confidence* bagi data kajicuaca

Bil	Set Item	<i>Support</i>	<i>Confidence</i>
1.	W1, H1	0.64	1.00
2.	W1, H6	0.00	0.00
3.	W2, H1	0.36	0.99
...
...
690.	R10, WS9	0.00	0.00

Nilai *support* dan *confidence* ini akan digunakan untuk mengasingkan set item yang *frequent* dan set item yang tidak *frequent*. Ia ditentukan dengan membandingkan nilai *support* dan *confidence* yang diperolehi dengan nilai minimum *support* dan juga nilai minimum *confidence*. Nilai minimum *support* di dalam kajian ini ialah 0.1 manakala nilai minimum *confidence* pula ialah 0.1. Selepas dibuat perbandingan di antara nilai *support* dan *confidence* dengan nilai minimum masing-masing, set item yang *frequent* akan diperolehi. Kemudian, set item yang *frequent* ini dikelompokkan dan ia menjadi data input kepada proses peramalan taburan hujan. Manakala set item yang tidak *frequent* tidak digunakan bagi tujuan pengujian peramalan taburan hujan di dalam kajian ini. Jadual 5 di bawah menunjukkan sebahagian daripada set item yang *frequent*. Senarai set item *frequent* yang diperolehi pula boleh dirujuk di Lampiran D.

Jadual 5 : Kelompok set item yang *frequent*

Kelompok	Set Item
Kelompok 1	W1, H1
Kelompok 2	W2, H1
Kelompok 3	W1, E1
...	...
..	...
Kelompok 47	R2, WS1

Bagi tujuan pengujian kelompok data kajicuaca manakah yang lebih baik, beberapa eksperimen pengujian telah dilakukan. Pengujian ini dilakukan dengan menggunakan kelompok-kelompok data kajicuaca yang dihasilkan, iaitu kelompok data kajicuaca yang sama dan juga kelompok data kajicuaca yang berlainan. Kelompok-kelompok ini digunakan sebagai data input kepada pakej NeuNetPro bagi melakukan peramalan taburan hujan. Di dalam kajian ini, pelaksanaan eksperimen peramalan taburan hujan boleh dibahagikan kepada dua bahagian iaitu peramalan menggunakan kelompok data kajicuaca yang sama dan peramalan menggunakan kelompok data kajicuaca yang berlainan.

Bagi tujuan peramalan taburan hujan dengan menggunakan atribut daripada kelompok data kajicuaca yang sama sebagai data input, sebanyak tiga eksperimen yang berkaitan telah dijalankan, iaitu;

- Eksperimen A - melibatkan atribut kelembapan dan suhu, iaitu (H1, T3).
- Eksperimen B - melibatkan atribut kelajuan angin dan tekanan, iaitu (WS1, TE1).
- Eksperimen C - melibatkan atribut windvane dan energy, iaitu (W2, E1).

Jadual 6(a), 6(b) dan 6(c) berikut menunjukkan contoh hasil peramalan taburan hujan bagi ketiga-tiga eksperimen-eksperimen yang telah dilaksanakan.

Jadual 6(a): Keputusan peramalan taburan hujan bagi Eksperimen A

Tarikh	Masa	Taburan Hujan (mm ³)	Nilai yang Diramal (mm ³)	Perbezaan (mm ³)
8/1/00	11:00	0	0.22	0.22
8/1/00	17:00	1.0	0.18	0.18
8/1/00	18:00	0	-0.09	0.09
8/11/00	4:00	0	0.06	0.06
...
8/11/00	8:00	0	-0.02	0.02

Daripada jadual di atas, didapati perbezaan di antara nilai sebenar taburan hujan dan nilai yang diramal adalah tidak sekata.

Jadual 6(b): Keputusan peramalan taburan hujan bagi Eksperimen B

Tarikh	Masa	Taburan Hujan (mm ³)	Nilai yang Diramal (mm ³)	Perbezaan (mm ³)
8/1/00	0:00	0	0.02	0.02
8/1/00	1:00	0	0.02	0.02
8/1/00	2:00	0	-0.13	0.13
...
...
8/11/00	9:00	0	0.22	0.22

Jadual di atas menunjukkan keputusan peramalan taburan hujan yang menggunakan kelompok atribut data kajicuaca yang sama iaitu atribut kelajuan angin dan tekanan.

Jadual 6(c): Keputusan peramalan taburan hujan bagi Eksperimen C

Tarikh	Masa	Taburan Hujan (mm ³)	Nilai yang Diramal (mm ³)	Perbezaan (mm ³)
8/1/00	1:00	0	0.01	0.01
8/1/00	2:00	0	0.03	0.03
8/1/00	3:00	0	0.05	0.05
...
...
8/11/00	0:00	0	-0.03	0.03

Jadual 6(c) menunjukkan keputusan peramalan taburan hujan yang menggunakan kelompok atribut data kajicuaca yang sama bagi eksperimen C, di mana ia melibatkan atribut *windvane* dan *energy*.

Bagi peramalan taburan hujan dengan menggunakan atribut daripada kelompok data kajicuaca yang berlainan sebagai data input pula, sebanyak dua eksperimen yang berkaitan telah dijalankan, iaitu;

- a) Eksperimen D - melibatkan atribut dari kelompok yang berlainan iaitu *windvane* dan suhu (W2, T3).

- b) Eksperimen E - melibatkan atribut dari kelompok berlainan, iaitu atribut *energy* dan suhu (E1,T4).

Jadual 7(a) dan 7(b) di bawah menunjukkan hasil peramalan taburan hujan bagi ketiga-tiga eksperimen-eksperimen yang telah dilakukan.

Jadual 7(d): Keputusan peramalan taburan hujan bagi Eksperimen D

Tarikh	Masa	Taburan Hujan (mm ³)	Nilai yang Diramal (mm ³)	Perbezaan (mm ³)
8/1/00	11:00	0	0.27	0.27
8/1/00	17:00	1	0.80	0.20
8/3/00	3:00	0	0.03	0.03
...
8/11/00	7:00	0	-0.07	0.07

Jadual di atas menunjukkan hasil peramalan taburan hujan menggunakan kelompok atribut data kajicuaca yang berlainan, iaitu atribut *windvane* dan suhu.

Jadual 7(e): Keputusan peramalan taburan hujan bagi Eksperimen E

Tarikh	Masa	Taburan Hujan (mm ³)	Nilai yang Diramal (mm ³)	Perbezaan (mm ³)
8/1/00	12:00	0	1.37	1.37
8/1/00	16:00	1.5	0.94	0.56
8/3/00	11:00	5	1.35	3.65
...
8/10/00	15:00	0.5	0.78	0.28

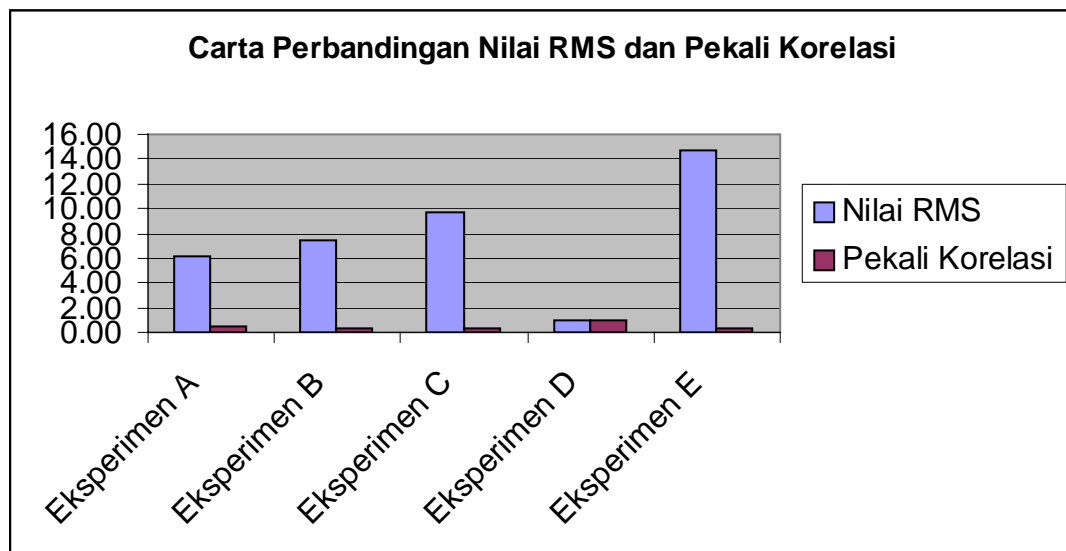
Jadual 7(e) di atas menunjukkan keputusan peramalan taburan hujan yang menggunakan kelompok atribut data kajicuaca yang berlainan, di mana atribut *energy* dan suhu telah digunakan sebagai data input kepada proses peramalan tersebut.

Berdasarkan kepada lima eksperimen yang telah dijalankan, penganalisan dan perbandingan telah dibuat ke atas hasil peramalan yang dilakukan terhadap data kajicuaca yang telah dikelompokkan dengan menggunakan teknik peraturan kesatuan. Tujuan perbandingan ini dilakukan ialah untuk melihat atau menilai sejauhmanakah prestasi hasil peramalan taburan hujan yang dikelompokkan dengan menggunakan teknik peraturan kesatuan dan juga kesan penggunaan kelompok data kajicuaca yang sama dan juga kelompok data kajicuaca yang berlainan ke atas hasil peramalan taburan hujan. Di dalam kajian ini, pengukuran prestasi peramalan taburan hujan yang telah dihasilkan dibuat berdasarkan kepada dua faktor iaitu nilai ralat min punca kuasa dua (RMS) dan juga nilai pekali korelasi yang diperolehi daripada kelima-lima eksperimen yang telah dilakukan. Jadual 8 di bawah menunjukkan keputusan nilai RMS dan juga nilai pekali korelasi tersebut.

Jadual 8 : Keputusan nilai RMS dan pekali korelasi

Eksperimen	Nilai RMS	Pekali Korelasi
Eksperimen A	6.15	0.52
Eksperimen B	7.46	0.38
Eksperimen C	9.65	0.39
Eksperimen D	1.02	0.91
Eksperimen E	14.71	0.29

Rajah 2 di bawah pula menunjukkan perbandingan di antara keputusan pengukuran peramalan yang diperolehi.



Rajah 2 : Perbandingan keputusan peramalan bagi setiap eksperimen

Rajah di atas menunjukkan nilai RMS yang dihasilkan di dalam eksperimen E jauh lebih tinggi berbanding dengan eksperimen-eksperimen yang lain. Manakala nilai pekali korelasi dalam eksperimen D pula memberikan nilai yang tertinggi di mana ia menghampiri nilai 1. Ini menunjukkan bahawa prestasi peramalan yang dihasilkan oleh eksperimen D merupakan yang terbaik jika dibandingkan dengan eksperimen yang lain. Oleh yang demikian, dapatlah dirumuskan bahawa penggunaan atribut di dalam kelompok yang berlainan menghasilkan keputusan peramalan taburan hujan yang lebih tepat berbanding dengan penggunaan kelompok yang sama. Ini boleh dibuktikan daripada nilai pekali korelasi di dalam eksperimen D yang menghampiri nilai 1 berbanding dengan eksperimen peramalan taburan hujan menggunakan kelompok yang sama sebagai data input kepada peramalan.

6.0 Perbincangan

Di dalam kajian ini, eksperimen peramalan taburan hujan yang telah dijalankan adalah bertujuan untuk melihat keupayaan teknik pengelompokan peraturan kesatuan di dalam memberi nilai ramalan yang tepat ke atas data taburan hujan. Hasil daripada eksperimen yang telah dijalankan menunjukkan bahawa teknik pengelompokan peraturan kesatuan memberikan prestasi peramalan yang baik. Selain daripada itu, kajian juga menunjukkan bahawa penggunaan kelompok data kajicuaca yang berlainan sebagai data input kepada proses

peramalan taburan hujan juga memberikan prestasi peramalan yang lebih baik berbanding dengan kelompok data kajicuaca yang sama.

Oleh yang demikian, maka dapatlah dirumuskan bahawa teknik pengelompokan peraturan kesatuan dapat meningkatkan prestasi peramalan taburan hujan. Di samping itu juga faktor kelompok yang digunakan sebagai data input kepada proses peramalan turut mempengaruhi keputusan peramalan, di mana penggunaan kelompok yang berlainan dapat memberikan prestasi peramalan yang lebih baik berbanding dengan penggunaan kelompok yang sama.

Terdapat beberapa masalah yang telah dihadapi di dalam melaksanakan kajian ini, di antaranya ialah;

- i) Masalah yang dihadapi daripada data-data kajicuaca yang diperolehi, di mana terdapat sebahagian data-data kajicuaca tersebut adalah data melampau. Ini telah menyebabkan keputusan analisa iaitu hasil nilai ralat dan peramalan taburan hujan kurang memuaskan. Data melampau bermaksud julat antara data yang berturutan adalah bersaiz besar.
- ii) Bagi tujuan pengujian peramalan taburan hujan, kajian ini telah menggunakan pakej NeuNet Pro. Pakej ini mempunyai kelemahan yang tersendiri di mana ia tidak boleh menjanakan proses peramalan jika bilangan rekod data input yang digunakan adalah kurang daripada 10. Selain daripada itu, pakej ini juga tidak boleh digunakan untuk membuat peramalan masa hadapan.

Kekangan bagi kajian ini ialah set item yang tidak *frequent* tidak digunakan di dalam pengujian peramalan taburan hujan ini. Manakala anggapan yang dibuat di dalam kajian ini ialah jarak di antara atribut data kajicuaca yang paling kecil dianggap mempunyai ciri persamaan yang kuat dan begitulah sebaliknya. Selain daripada itu, data-data kajicuaca yang digunakan dianggap bersih dan bebas dari hingar.

Kajian terhadap teknik pengelompokan data kajicuaca ini merupakan satu bidang yang menarik untuk diselidiki. Ini kerana operasi pengelompokan telah banyak digunakan di dalam pelbagai bidang terutamanya bidang pengelompokan imej (Shahliza, 1999), aksara (Azah Kamilah, 1999) data tak berstruktur jenis teks (Siti Sakira, 1998) dan sebagainya. Dan terdapat banyak teknik pengelompokan terutamanya teknik perlombongan data yang boleh digunakan untuk melakukan pengelompokan.

Dalam kajian ini, penggunaan teknik peraturan kesatuan, iaitu salah satu teknik perlombongan data bagi melakukan operasi pengelompokan telah dilaksanakan. Daripada eksperimen yang telah dilakukan, kajian ini telah dapat memberikan satu cadangan atau alternatif baru kepada pihak JPKM untuk melakukan penganalisaan data kajicuaca bagi tujuan peramalan. Manakala kekurangan yang terdapat pada kajian ini, diharapkan ianya dapat dikembangkan lagi pada masa akan datang. Umpamanya;

- i) Menggunakan data kajicuaca dalam bentuk harian atau mingguan.
- ii) Menggunakan pakej yang lain bagi tujuan perbandingan peramalan, contohnya pakej NeuralWorks dan lain-lain.
- iii) Membuat peramalan taburan hujan pada masa hadapan.
- iv) Menggunakan teknik pengelompokan yang lain seperti teknik jiran terdekat dan statistik.

7.0 Kesimpulan

Kajian ini dijalankan untuk mengkaji teknik pengelompokan data kajicuaca bagi tujuan meramal taburan hujan. Terdapat pelbagai cara yang boleh digunakan untuk melakukan pengelompokan data kajicuaca, di antaranya ialah dengan menggunakan kaedah statistik,

teknik perlombongan data dan sebagainya. Sehubungan dengan itu, kajian ini dilaksanakan dengan menggunakan teknik peraturan kesatuan untuk mengelompokkan data kajicuaca bagi tujuan peramalan taburan hujan. Hasil kajian ini menunjukkan prestasi peramalan taburan hujan yang menggunakan kelompok atribut data kajicuaca yang berlainan adalah lebih baik berbanding dengan menggunakan kelompok atribut data kajicuaca yang sama.

RUJUKAN

- Arakawa, H. (1969). "Climates of Northern dan Eastern Asia." World Survey of Climatology, Vol.8, Amsterdam : Elsevier Publishing Company.
- Azah Kamilah Muda (1999). "Pengekstrakan Ciri dan Perangkaian Aksara Taip Pelbagai Variasi : Kajian Perbandingan Antara Teknik Momen Geometric Tak Varian, Momen Regular dan Momen Legendre." Universiti Teknologi Malaysia: Tesis Sarjana.
- Berson, A. dan Smith, S. J. (1997). "Data Warehousing, Data Mining dan OLAP." New York, USA : McGraw-Hill.
- Cheang, Boon Khean dan Ooi, See Hai (1989). "Review of the November 1987 – March 1988 Northeast Monsoon in Malaysia, Part I : Rainfall Patern." Nota Teknikal No: 36, April 1989. Perkhidmatan Kajicuaca Malaysia.
- Chen, Tao dan Takagi, Mikio (1993). "Rainfall Prediction of Geostationary Meteorological Satellite Images Using Artificial Neural Network."Jepun : University of Tokyo.
- Cheng, Shan (1998). "Statistical Approaches to Predictive Modeling in LargeDatabases." Simon Fraser University : Tesis Sarjana.
- Choenni, R. dan Siebes, R. (1996). "Data Mining Applications." Computer Aided Decision Support in Telecommunications. UK: Chapman dan Hall.
- Fact, E. W. (1987). "Quantitative Methods." London.
- Fayyad, U. (1997). "Data Mining and Knowledge Discovery in Databases : Implications for Scientific Databases." Proceedings of the IEEE.
- Fulton, R., J. Breidenbach, D. J. Seo, D. Miller, dan T. O'Bannon, (1998). "The WSR-88D rainfall algorithm." Weather and Forecasting, **13**, 377-395.
- Hua, Zhu (1998). "On-line Analytical Mining of Association Rules." Simon Fraser University : Tesis Sarjana.
- Hofmann, T. dan Buhmann, J. (1997). "Inferring Hierarchical Clustering Structures by Deterministic Annealing." [Online] Available : <http://www-dbv.cs.uni-bonn.de>
- Jabatan Perkhidmatan Kajicuaca Malaysia (2002). [Online] Available : http://www.kjc.gov.my/main_m.html
- Kartasapoetra, A.G. (1986). "Klimatologi Pengaruh Iklim Terhadap Tanah dan Tanaman." Jakarta : Penerbit Bina Aksara.
- Kaski, S. (1997). "Data Exploration Using Self-Organizing Maps." Helsinki University of Technology : Tesis Ph.D.,
- Korkmaz, E. (1997). " A Method For Improving Automatic Word Categorization." The Middle East Technical University : Tesis Sarjana.
- Liu, James N. K. dan Lee, Raymond S.T. (1999). "Rainfall Forecasting from Multiple Point Sources Using Neural Networks." Hong Kong : Hong Kong Polytechnic University.

- McCullagh, J., Bluff, K. dan Hendtlass, T. (1995). "A Neural Network Model for Rainfall Estimation." IEEE.
- McCullagh, J., Bluff, K. dan Hendtlass, T. (1999). "Evolving Expert Neural Networks for Meteorological Rainfall Estimation." IEEE.
- Ordóñez, C.; Omiecinski, E.; de Braal, L.; Santana, C.A.; Ezquerro, N.; Taboada, J.A.; Cooke, D.; Krawczynska, E.; Garcia, E.V. (2001). "Mining constrained association rules to predict heart disease." *Proceedings IEEE International Conference on Data Mining*: 433 –440.
- Ronkainen, P. (1998). "Attribute Similarity Event Sequence Similarity in Data Mining." Department of Computer Science : University of Helsinki.
- Rosdiah Dahlan (1999). "Analisa Taburan Hujan Menggunakan Model Transfer Function." Universiti Teknologi Malaysia : Tesis Sarjana Muda.
- Roselina Sallehuddin (1999). "Penggunaan Model Rangkaian Neural dalam Peramalan Siri Masa Bermusim." Universiti Teknologi Malaysia : Tesis Sarjana.
- Sham Sani (1984). "Beberapa Aspek Iklim Negeri Sabah." UKM : Bangi.
- Shahliza Abd Halim (1999). "Feature Extraction and Clustering of Trademark Images : A Comparison between Geometric Invariant and Zernike Moment Invariant." Universiti Teknologi Malaysia : Tesis Sarjana.
- Sarjon Defit dan Mohd Noor Mad Sap (2000). "Predictive Data Mining Based On Similarity and Clustering Methods." *Jurnal Teknologi Maklumat* , Fakulti Sains Komputer dan Sistem Maklumat : UTM. Jilid 12, Bil.2.
- Siti Sakira Kamarudin (1998). "Clustering Unstructured Data From A Text Database Using Self-Organizing Maps." Universiti Teknologi Malaysia : Tesis Sarjana.
- TwoCrows Corporation, (1999). "Introduction To Data Mining and Knowledge Discovery.", 3rd Edition : USA.
- Wang, W. (1999). "Predictive Modeling Based on Classification and Pattern Matching Methods." Simon Fraser University : Tesis Sarjana.
- Yair, E.Z., Dan, R. dan Dmitry, Z. (1999). "Word Prediction and Clustering." Dept. of Computer Science. University of Illinois : Urbana.
- Yeong, Yoke Wei (1996). "Penggunaan Model Siri Masa untuk Menganalisis dan Meramal Kadar Hujan." Universiti Teknologi Malaysia : Tesis Sarjana Muda.

LAMPIRAN A – SENARAI KELAS-KELAS

WINDVANE		RADIATION	
W1	0.0 - 200.0	R1	(-100.00 hingga 0.00)
W2	200.01 - 400.0	R2	0.1 - 150.0
W3	400.01 - 600.0	R3	151.1 - 300.0
W4	600.01 - 800.0	R4	300.1 - 450.0
W5	800.01 - 1000.0	R5	450.1 - 600.0
W6	1000.01 -1200.0	R6	600.1 - 750.0
W7	1200.01 - 1400.0	R7	750.1 - 900.0
W8	1400.01 - 1600.0	R8	900.1 - 1050.0
W9	1600.01 - 1800.0	R9	1050.1 - 1200.0
W10	1800.01 - 2000.0	R10	1200.1 - 1350.0
KELAJUAN ANGIN		TABURAN HUJAN	
WS1	0.00 - 0.10	RF1	0.0 - 5.0
WS2	0.11 - 0.20	RF2	5.1 - 10.0
WS3	0.21 - 0.30	RF3	10.1 - 15.0
WS4	0.31 - 0.40	RF4	15.1 - 20.0
WS5	0.41 - 0.50	RF5	20.1 - 25.0
WS6	0.51 - 0.60	RF6	25.1 - 30.0
WS7	0.61 - 0.70	RF7	30.1 - 35.0
WS8	0.71 - 0.80	RF8	35.1 - 40.0
WS9	0.81 - 0.90	RF9	40.1 - 45.0
WS10	0.91 - 1.00	RF10	45.1 - 50.0
KELEMBAPAN		TEKANAN	
H1	0.0 - 12.5	TE1	(-10.0 hingga 86.0)
H2	12.6 - 25.0	TE2	86.1 - 182.0
H3	25.1 - 37.5	TE3	182.1 - 278.0
H4	37.6 - 50.0	TE4	278.1 - 374.0
H5	50.1 - 62.5	TE5	374.1 - 470.0
H6	62.6 - 75.0	TE6	470.1 - 566.0
		TE7	566.1 - 662.0
		TE8	662.1 - 758.0
		TE9	758.1 - 854.0
		TE10	854.1 - 950.0
ENERGY		SUHU	
E1	(-1.0 - 99.00)	T1	(-13.0 hingga -8.0)
E2	99.01 - 199.00	T2	(-7.9 hingga -3.0)
E3	199.01 -299.00	T3	(-2.9 hingga 2.0)
E4	299.01 – 399.00	T4	2.1 - 7.0
E5	399.01 – 499.00	T5	7.1 - 12.0
E6	499.01 – 599.00	T6	12.1 - 17.0
E7	599.01 – 699.00	T7	17.1 - 22.0
E8	699.01 – 799.00	T8	22.1 - 27.0